



# A model of the ACE2 structure and function as a SARS-CoV receptor

Ponraj Prabakaran, Xiaodong Xiao, and Dimiter S. Dimitrov\*

Laboratory of Experimental and Computational Biology, CCR, NCI-Frederick, NIH, Frederick, MD 21702-1201, USA

Received 25 November 2003

## Abstract

The angiotensin-converting enzyme 2 (ACE2) is an important regulator of the renin–angiotensin system and was very recently identified as a functional receptor for the SARS virus. The ACE2 sequence is similar (sequence identities 43% and 35%, and similarities 61% and 55%, respectively) to those of the testis-specific form of ACE (tACE) and the *Drosophila* homolog of ACE (AnCE). The high level of sequence similarity allowed us to build a robust homology model of the ACE2 structure with a root-mean-square deviation from the aligned crystal structures of tACE and AnCE less than 0.5 Å. A prominent feature of the model is a deep channel on the top of the molecule that contains the catalytic site. Negatively charged ridges surrounding the channel may provide a possible binding site for the positively charged receptor-binding domain (RBD) of the *S*-glycoprotein, which we recently identified [Biochem. Biophys. Res. Commun. 312 (2003) 1159]. Several distinct patches of hydrophobic residues at the ACE2 surface were noted at close proximity to the charged ridges that could contribute to binding. These results suggest a possible binding region for the SARS-CoV *S*-glycoprotein on ACE2 and could help in the design of experiments to further elucidate the structure and function of ACE2.

© 2003 Elsevier Inc. All rights reserved.

**Keywords:** ACE2; Structure prediction; SARS-CoV; *S*-Glycoprotein; Receptor; Binding

Enveloped viruses enter cells by binding their envelope glycoproteins to cell surface receptors followed by conformational changes leading to membrane fusion and delivery of the genome in the cytoplasm [1]. Very recently, the angiotensin-converting enzyme 2 (ACE2) was identified as a functional receptor for the SARS coronavirus (SARS-CoV) [2] and its binding site on the SARS-CoV *S*-glycoprotein was localized between amino acid residues 303 and 537 [3]. ACE2 is a homolog of the metalloprotease angiotensin-converting enzyme ACE [4,5] and was found to be an essential regulator of heart function [6]. ACE exists in two isoforms—somatic ACE, which has two homologous domains each containing an active catalytic site, and testis-specific ACE (tACE), which corresponds to the C domain of somatic ACE and has only one active site. ACE2 has a high level of similarity (sequence identities 43% and 35%, and similarities 61% and 55%, respectively) to tACE and the *Drosophila* homolog of ACE (AnCE). Recently, the crystal struc-

tures of tACE [7] and the *Drosophila* ACE homolog AnCE [8] have been determined at resolutions 2.0 and 2.4 Å, respectively.

These crystal structures were used as templates to build an accurate (root-mean-square deviation (rmsd) less than 0.5 Å) three-dimensional (3D) model of ACE2 by comparative (homology) modeling. Based on the ACE2 model, an analysis of the receptor-binding domain (RBD) of the SARS-CoV *S*-glycoprotein, and similarity with other interactions of viral envelope glycoproteins (Env) with receptors [9], we propose a possible mechanism of the ACE2 function as a receptor for the SARS virus. The analysis of the ACE2 model could also help in the design of experiments to further elucidate the structure and the dual function of ACE2.

## Materials and methods

**Homology modeling of the ACE2 structure.** The sequences of ACE2, tACE, and AnCE were aligned by using the multiple sequence alignment program CLUSTALW [10]. The comparative modeling procedure COMPOSER [11,12] implemented in SYBYL6.9 (Tripos, St.

\* Corresponding author. Fax: 1-301-846-5598.

E-mail address: [dimitrov@ncifcrf.gov](mailto:dimitrov@ncifcrf.gov) (D.S. Dimitrov).

Louis, MO) was used to build a 3D model of the ACE2 structure. We used the tACE and AnCE structures to find out topologically equivalent residues based on structural alignment and the structurally conserved regions (SCRs) were modeled. The structurally variable regions (loops) were modeled by using loops either from the corresponding location of the homologous protein or from the general protein database. The 3D model of ACE2 was then subjected to energy minimization by using standard Tripos force fields and finally validated with the PROCHECK program [13]. The coordinates were deposited to the protein data bank (PDB) (code: 1RIX).

**Modeling of a fragment containing the S-glycoprotein RBD.** The threading was performed by the GeneFold module in SYBYL6.9 that uses three different scoring functions combining sequence and structure information [14]. The cytokine binding region of gp130 (PDB code 1BQU) was selected as a putative target structure with high scores from all the three scoring algorithms. The loops in the model were generated using protein loop search function of the GeneFold by scanning protein structural database for similar sequences. Among seven cysteine residues in the model, three pairs of disulfide bonds (S–S) were simulated and the resulting model was energy minimized. The model was validated with the PROCHECK program [13].

**Modeling of glycosylation sites, electrostatic analysis, solvent accessibility, and surface hydrophobicity.** The sequences of both ACE2 and the S RBD were scanned against the PROSITE [15] motifs in order to locate potential glycosylation sites. Six N-glycosylation sites with high probability of occurrences on ACE2 were predicted by PROSITE. Fully surface-exposed asparagine (N) residues were found at five of these sites, which were modeled by attaching N-acetylglucosamine moieties. Three N-glycosylation sites were found in the S RBD fragment and were modeled similarly. The areas of solvent accessibility (ASA) were calculated with a probe radius of 1.4 Å by using the Lee and Richards' algorithm [16]. Electrostatic potentials were calculated by using the program GRASP [17] with the following parameters: a protein dielectric constant of 2.0, a solvent dielectric constant of 80, an ion exclusion radius of 2.0 Å, a probe radius of 1.4 Å, and an ionic strength of 0.14 M. The calculated potentials were displayed at the solvent-accessible surface. The visualization of solvent accessibility, super-positioning of molecules, and calculation of surface hydrophobicity were performed by using InsightII. The hydrophobicity of the surface residues was calculated according to the Kyte–Doolittle method [18] with a window size of 5, and hydrophobic and hydrophilic levels of 0.7 and –2.4, respectively.

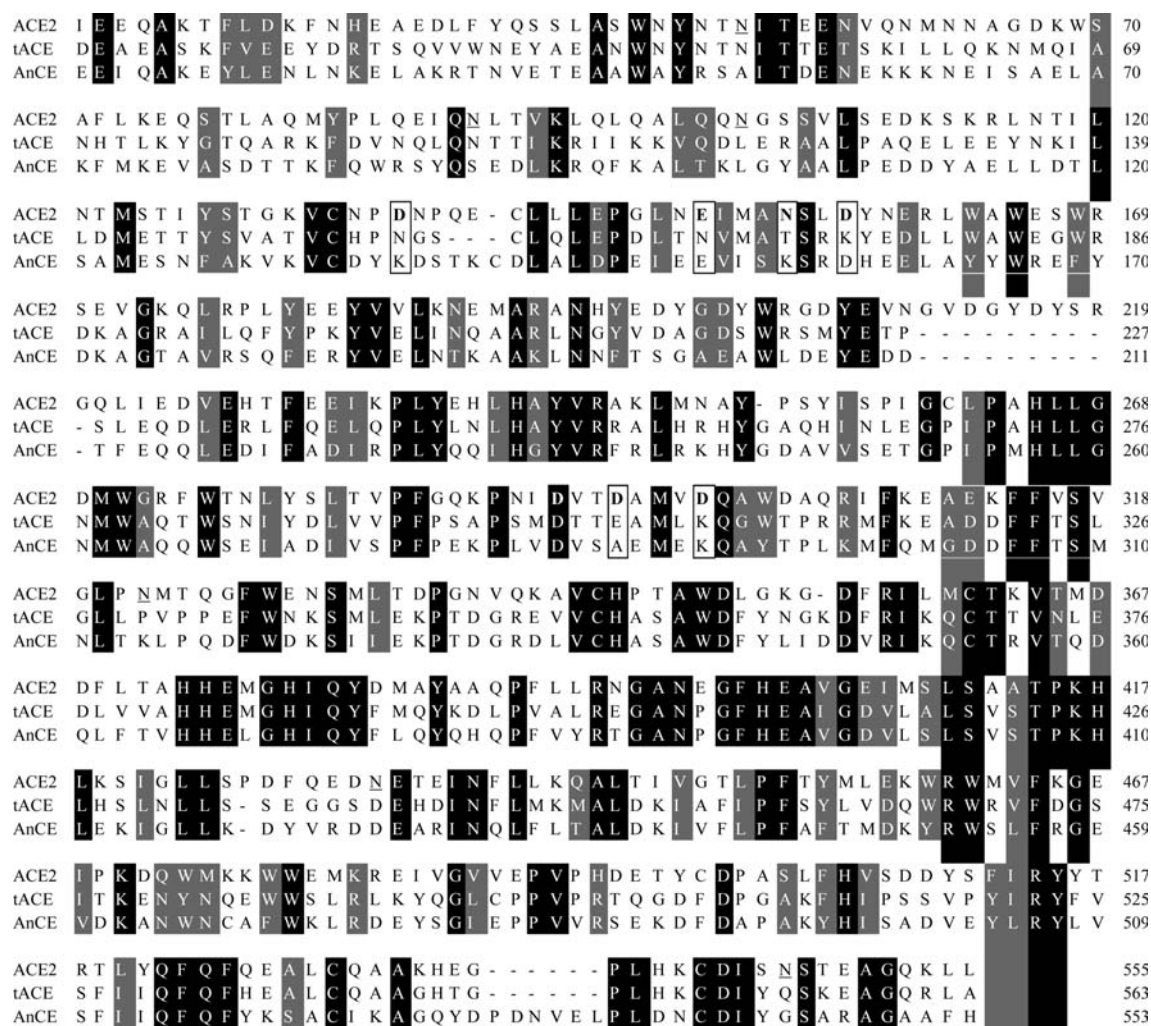


Fig. 1. Multiple sequence alignment of ACE2, tACE, and AnCE using CLUSTALW. The sequence numbering is the same as in the crystal structures. The N-glycosylation sites are underlined; putative binding residues in ACE2 are in boldface letters and boxed along with the corresponding aligned residues in tACE and AnCE. The identical and similar residues are shown in black and gray backgrounds, respectively.

## Results and discussion

### The ACE2 model

To begin to understand the interactions between the SARS-CoV *S*-glycoprotein and its recently identified receptor, ACE2, we attempted to develop an accurate model of the ACE2 3D structure. Currently, the only computational methodology that allows prediction of protein 3D structures with high accuracy (an rms error lower than 2 Å) is comparative (homology) modeling. However, it requires sequence identity of the target protein with templates of known 3D structures higher than about 30% for accurate structure prediction. We found two proteins, tACE and AnCE, with available high resolution crystal structures, and ACE2 sequence identities of 43% and 35% (sequence similarities are 61% and 55%), respectively; the sequence alignment of ACE2 with tACE and AnCE2 is shown in Fig. 1. Therefore, we have used homology modeling to build an accurate 3D model of ACE2 as described in Materials and methods.

The architecture of the ACE2 model is very similar to the crystal structure of tACE (Fig. 2A). The superposition of the ACE2 model structure with the template structures of tACE and AnCE (Fig. 2B) shows very small deviation (rmsd less than 0.5 Å). A major feature of the ACE2 structure (and the template structures) is a deep channel on the top of the molecule that contains the catalytic site (Fig. 3A). A comprehensive analysis of the structure and function of the catalytic site was very recently reported after our model was completed [19]; here we will not discuss the enzymatic function of ACE2 but rather use the enzymatic site location for reference purposes. The channel is surrounded by ridges containing loops, helices, and a portion of a  $\beta$ -sheet. The

long loop between N210 and Q221 that is missing in tACE and AnCE (Fig. 1) is on the ACE2 surface (Fig. 2B); note that the orientation of ACE2 in Fig. 2B is different than in Fig. 2A in order to show this loop. Potential N-glycosylation sites were identified at six positions: 53, 90, 103, 322, 432, and 546, but only two of them (53 and 90) were aligned with the tACE structure (Fig. 1). They shared the pattern NXTX (except 103) and were modeled with a *N*-acetylglucosamine moiety (Fig. 3B). The direction of the main chain is illustrated in Fig. 3C.

### ACE2 surface potential, solvent accessibility, hydrophobicity analysis, and carbohydrate distribution

Interactions of viral attachment proteins with protein receptor molecules are mostly determined by complementarity in surface charge distribution, hydrophobic interactions, and geometry; typically carbohydrates are excluded from the binding sites [9]. In an attempt to provide working hypothesis for possible regions involved in the interaction of the *S*-glycoprotein with its receptor we analyzed the ACE2 surface potential, solvent accessibility, hydrophobicity, and carbohydrate distribution. The surface of the deep channel at the top of ACE2 and the surrounding ridges is highly negatively charged (Fig. 3A). These ridges contain residues D136, E150, N154, D157, D292, D295, and D299 some of which have large ASA values, e.g., D136, N154, and D157 have 109, 108, and 80 Å<sup>2</sup>, respectively (Fig. 4). Comparison of these residues with the corresponding residues from ACE that do not support fusion mediated by the *S*-glycoprotein [2], and mouse ACE2 that binds to *S* but with somewhat lower affinity than human ACE2 (M. Farzan, personal communication) (Fig. 5)

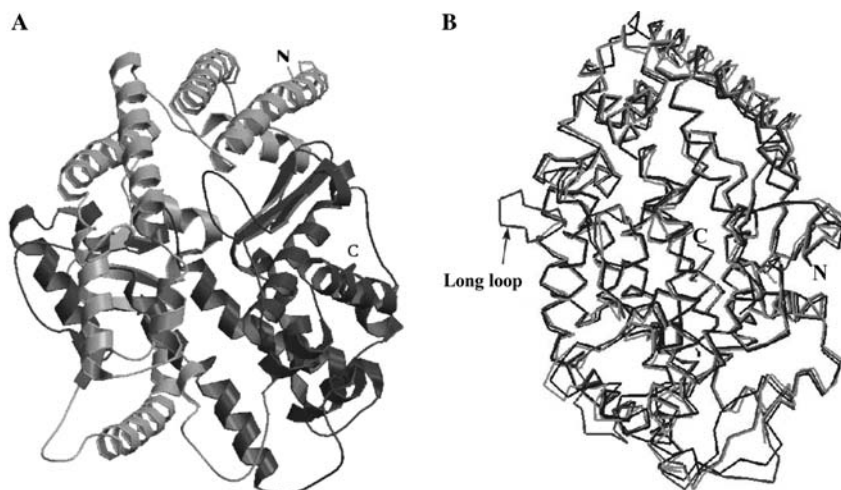


Fig. 2. The model of the ACE2 structure. (A) A ribbon representation of the ACE2 model. The N- and C-termini are indicated. (B) Superposition of the ACE2 model structure with the crystal structures of tACE and AnCE based on the  $\alpha$ -atoms of ACE2, tACE, and AnCE (ACE2, dark gray; tACE, light gray; and AnCE, black). The long loop inserted between N210 and Q221 that is unique for ACE2 is indicated.

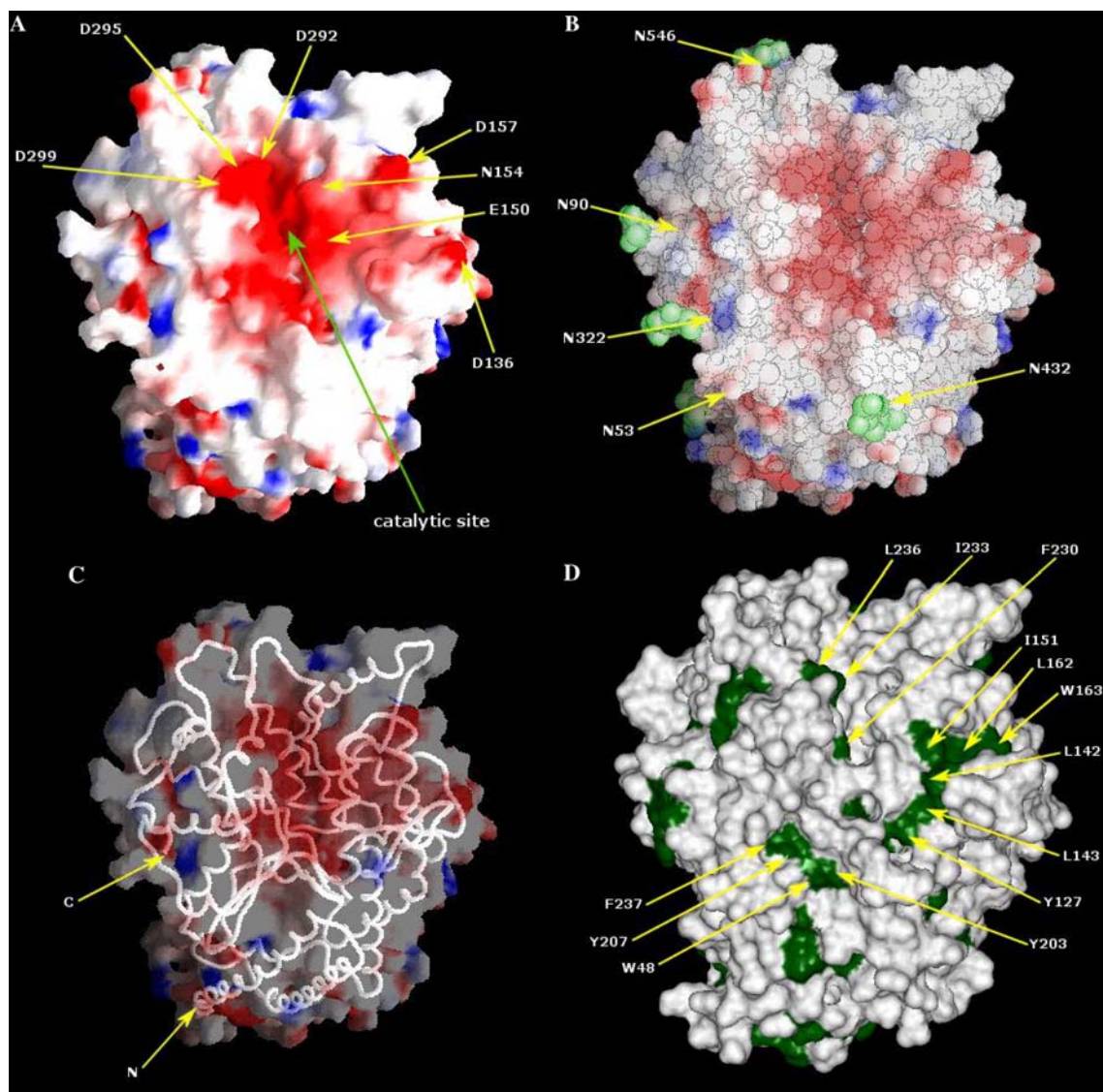


Fig. 3. Analysis of the ACE2 model structure. (A) Representation of negative and positive molecular surface potentials in red and blue, respectively. The channel at the top of the molecule containing the catalytic site and the surrounding ridges containing negatively charged residues is indicated. (B) Distribution of glycosylation sites (green) on the ACE2 surface. (C) The backbone wire representing the orientation of the main chain. (D) Distribution of hydrophobic patches on the ACE2 surface.

supports the possibility that some of these residues contribute to specific binding. The hydrophobicity analysis revealed distinct hydrophobic patches in close proximity to the negatively charged ridges (Fig. 3D). There are at least three hydrophobic regions comprising different residues including Phe, Trp, and Tyr which could contribute to binding in addition to the charged binding surface. All carbohydrate sites are topologically apart from the electronegative surface at the top of the molecule (Fig. 3B).

#### *Proposition of receptor–S-glycoprotein binding interactions*

The sequence similarity of the S-glycoprotein from the SARS virus with S-glycoproteins from other coro-

naviruses or other proteins whose structures are available in the PDB is about 20% or lower. The sequence similarity of the attachment glycoprotein (S1) from the SARS-CoV to other coronavirus S1 glycoproteins or other proteins with known 3D structures is even lower. Such low sequence similarity does not allow accurate homology modeling. Due to the absence of significant sequence similarity, we built a model by threading (data not shown) of a fragment (amino acid residues 300–537) containing the S RBD that we have recently identified [3]. The electrostatic analysis of the model revealed mostly positive charges on the surface and, particularly, an electronegative loop containing residues K439, R441, R444, H445, and K447. The hydrophobic analysis suggested several patches of hydrophobic residues around the positively charged loop region. One must note that

LYS 26 133.75	PRO 138 111.19	SER 257 57.50	SER 420 89.94
THR 27 106.68	GLN 139 124.05	PRO 258 65.89	GLY 422 45.02
ASP 30 90.32	GLU 140 96.63	ILE 259 167.18	PRO 426 120.83
LYS 31 146.81	PRO 146 62.39	SER 280 86.96	ASP 427 102.68
HIS 34 103.48	<b>GLU 150 82.31</b>	GLY 286 49.20	PHE 428 152.59
GLU 35 67.77	<b>ASN 154 107.65</b>	GLN 287 160.01	GLN 429 82.31
GLN 42 88.04	LEU 156 96.54	PRO 289 123.13	GLU 430 117.90
ASN 49 72.35	<b>ASP 157 79.60</b>	ASN 290 84.88	ASP 431 138.66
GLU 56 152.15	ASN 159 97.55	<b>ASP 292 83.20</b>	GLU 433 74.27
GLU 57 139.26	GLU 160 83.50	THR 294 52.85	THR 434 58.21
GLN 60 128.47	GLU 171 80.80	<b>ASP 295 54.90</b>	ILE 436 80.00
ASN 61 79.64	GLN 175 91.63	VAL 298 87.15	LEU 439 71.32
ASN 64 90.94	PRO 178 89.84	<b>ASP 299 95.50</b>	GLY 466 71.17
ASP 67 53.92	VAL 185 70.22	GLN 300 96.98	PRO 469 71.33
LYS 68 109.99	ARG 192 132.40	ALA 301 90.36	LYS 470 122.67
ALA 71 60.43	HIS 195 168.85	ASP 303 76.71	ASP 471 119.93
GLU 75 103.10	GLU 197 109.67	GLN 305 97.67	GLN 472 69.35
THR 78 67.46	ASP 206 71.43	ARG 306 96.24	LYS 475 132.50
LEU 79 73.79	GLU 208 146.30	LYS 309 81.39	GLU 479 81.98
GLN 81 110.23	VAL 209 76.31	LYS 313 119.53	ARG 482 104.76
MET 82 127.23	ASN 210 167.13	GLY 319 53.77	GLU 483 117.49
PRO 84 74.73	TYR 217 143.76	ASN 322 71.65	PRO 492 113.80
GLN 86 139.19	SER 218 82.84	THR 324 74.83	ASP 494 84.95
GLU 87 128.66	ARG 219 187.61	GLN 325 122.91	GLU 495 138.47
GLN 89 148.33	GLN 221 69.68	GLU 329 129.15	THR 496 85.99
ASN 90 69.13	GLU 224 73.82	ASN 330 68.41	TYR 510 93.92
LEU 91 131.77	ASP 225 75.54	THR 334 75.00	GLN 524 101.90
THR 92 69.98	HIS 228 88.24	ASP 335 76.53	GLU 527 95.03
LEU 95 83.29	GLU 231 81.03	GLY 337 84.09	ALA 528 47.83
GLN 98 93.66	GLU 232 121.46	ASN 338 115.17	GLN 531 138.22
SER 105 45.43	LYS 234 92.20	VAL 339 160.63	LYS 534 157.11
SER 109 72.60	PRO 235 74.61	LYS 353 109.79	GLU 536 137.88
GLU 110 129.44	LEU 236 88.19	ASP 367 76.79	GLY 537 61.61
ASP 111 86.02	GLU 238 84.35	ALA 386 39.72	PRO 538 67.51
LYS 114 116.29	HIS 239 127.38	ALA 387 103.41	LEU 539 75.65
ARG 115 111.85	ALA 242 46.26	GLN 388 77.36	LYS 541 93.81
THR 118 54.89	ALA 246 55.51	ASN 394 96.86	ASN 546 96.10
THR 129 63.36	ASN 250 117.89	ALA 396 81.30	THR 548 84.73
PRO 135 145.59	PRO 253 91.46	PRO 415 49.57	GLU 549 96.53
<b>ASP 136 108.67</b>	SER 254 80.22	LYS 416 129.10	GLN 552 114.05
ASN 137 58.91	TYR 255 141.34	LYS 419 93.66	LYS 553 118.29

Fig. 4. Solvent accessible surface areas (right column, in Å<sup>2</sup>) for ACE2 amino acid residues that are significantly exposed to solvent at the surface of the molecule. The cut-off for significant surface exposure here is assumed to be 45% ratio value defined as the ratio of side-chain surface area to a “random coil” value per residue in the tripeptide Gly–X–Gly. The middle column represents the amino acid residue number.

	155	167	171	174	300	303	307
Human ACE	ASN-ASN-THR-LYS-ASP-GLU-LYS						
	:	.	.	.		:	.
Human ACE2	ASP-GLU-ASN-ASP-ASP-ASP-ASP						
	.		.				:
Mouse ACE2	LYS-GLU-THR-ASP-ASP-ASP-ASN						
	136	150	154	157	292	295	299

Fig. 5. Conservation of amino acid residues in human ACE2, human ACE, and mouse ACE2 that could contribute to interactions with the S-glycoprotein. Identities are marked by a pipe (|), highly conservative replacements by a colon (:), and replacements with lower scores by a dot (.). The numbers denote the amino acid residue positions in the sequence. Note that the similarity of these ACE2 residues with the corresponding residues of mouse ACE2 is much higher than with the respective human ACE residues.

although the size of the fragment is relatively small, the S RBD modeling has limitations in the absence of a template structure(s) with high sequence identity. Thus, the RBD model could significantly deviate or even be completely different from the real structure. In this aspect modeling of the much larger S1 and S2 units is even less reliable. For example, a recent model [20] of S1 and S2 proposed putative receptor (thought to be CD13) binding regions located in S2 instead of S1 where RBDs of coronaviruses should be. This is why we used our RBD model mostly as an illustration of possible complementary charged surfaces, hydrophobic patches, and  $\beta$ -sheets, and complemented it with an analysis of the

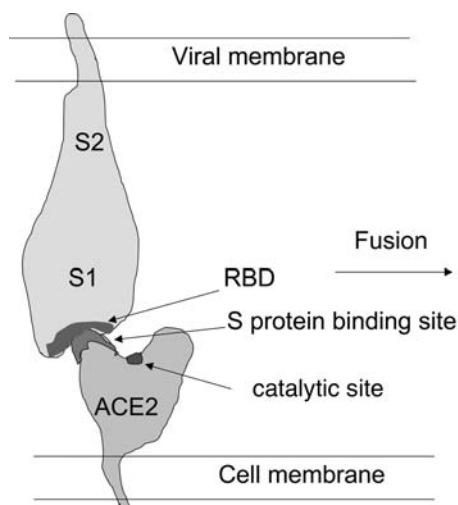


Fig. 6. Schematic representation of the interaction between ACE2 and the SARS-CoV S-glycoprotein leading to binding and fusion. The RBD is depicted as a surface containing a cavity(s) that binds a ridge(s) close to the deep channel containing the catalytic site.

secondary structure of the RBD fragment that also revealed predominance of  $\beta$ -sheets (data not shown). In progress are our experiments for the SARS-CoV S-glycoprotein RBD crystallization and determination of its 3D structure.

Typically virus receptors contain ridges that bind to cavities or to structures containing loops, cavities, and channels in the proteins mediating entry [9]. The model structure of ACE2 indicates that some or most of the ridges surrounding the cavity at the top of the molecule (Fig. 6) could serve as a likely binding region for the S-glycoprotein for the following reasons. First, the top of the molecule is far away of the membrane and is likely to be easier to reach than membrane proximal regions. Second, protruding structures are likely to be used for binding by viral proteins; it will also ensure geometric complementarity with concave surfaces as the S RBD domain could be based on our illustrative model (Fig. 6). Third, the negative charges of the ridges complement the positive charges of the RBD. Fourth, the hydrophobic patches around the charges could contribute to binding. Finally, the lack of carbohydrates at the top of the molecule could ensure high-affinity binding. Experiments currently in progress will determine the specific amino acid residues and their relative contribution to the interaction of ACE2 with the S-glycoprotein. The ACE2 model developed here, and this proposition of binding regions could help in the design and analysis of the experimental data, and the virus binding function of ACE2.

## Acknowledgments

We thank M. Farzan for communicating unpublished data and interesting discussions, S. Ravichandran for helpful discussions, and

the Advanced Biomedical Computing Center (ABCC), NCI-Frederick, for computing facilities.

## References

- [1] D.S. Dimitrov, Cell biology of virus entry, *Cell* 101 (2000) 697–702.
- [2] W. Li, M.J. Moore, N. Vasilieva, J. Sui, S.K. Wong, M.A. Berne, M. Somasundaran, J.L. Sullivan, K. Luzuriaga, T.C. Greenough, H. Choe, M. Farzan, Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus, *Nature* 426 (2003) 450–454.
- [3] X. Xiao, S. Chakraborti, A.S. Dimitrov, K. Gramatikoff, D.S. Dimitrov, The SARS-CoV S-glycoprotein: expression and functional characterization, *Biochem. Biophys. Res. Commun.* 312 (2003) 1159–1164.
- [4] M. Donoghue, F. Hsieh, E. Baronas, K. Godbout, M. Gosselin, N. Stagliano, M. Donovan, B. Woolf, K. Robison, R. Jeyaseelan, R.E. Breitbart, S. Acton, A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1–9, *Circ. Res.* 87 (2000) E1–E9.
- [5] S.R. Tipnis, N.M. Hooper, R. Hyde, E. Karran, G. Christie, A.J. Turner, A human homolog of angiotensin-converting enzyme. Cloning and functional expression as a captopril-insensitive carboxypeptidase, *J. Biol. Chem.* 275 (2000) 33238–33243.
- [6] M.A. Crackower, R. Sarao, G.Y. Oudit, C. Yagil, I. Kozieradzki, S.E. Scanga, A.J. Oliveira-dos-Santos, J. da Costa, L. Zhang, Y. Pei, J. Scholey, C.M. Ferrario, A.S. Manoukian, M.C. Chappell, P.H. Backx, Y. Yagil, J.M. Penninger, Angiotensin-converting enzyme 2 is an essential regulator of heart function, *Nature* 417 (2002) 822–828.
- [7] R. Natesh, S.L. Schwager, E.D. Sturrock, K.R. Acharya, Crystal structure of the human angiotensin-converting enzyme–lisinopril complex, *Nature* 421 (2003) 551–554.
- [8] H.M. Kim, D.R. Shin, O.J. Yoo, H. Lee, J.O. Lee, Crystal structure of *Drosophila* angiotensin I-converting enzyme bound to captopril and lisinopril, *FEBS Lett.* 538 (2003) 65–70.
- [9] D.S. Dimitrov, Virus entry: molecular mechanisms and biomedical applications, *Nat. Rev. Microbiol.*, 2004 (in press).
- [10] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673–4680.
- [11] M.J. Sutcliffe, F.R. Hayes, T.L. Blundell, Knowledge based modelling of homologous proteins, Part II: Rules for the conformations of substituted sidechains, *Protein Eng.* 1 (1987) 385–392.
- [12] M.J. Sutcliffe, I. Haneef, D. Carney, T.L. Blundell, Knowledge based modelling of homologous proteins, Part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures, *Protein Eng.* 1 (1987) 377–384.
- [13] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Cryst.* 26 (1993) 283–291.
- [14] L. Jaroszewski, L. Rychlewski, B. Zhang, A. Godzik, Fold prediction by a hierarchy of sequence, threading, and modeling methods, *Protein Sci.* 7 (1998) 1431–1440.
- [15] C.J. Sigrist, L. Cerutti, N. Hulo, A. Gattiker, L. Falquet, M. Pagni, A. Bairoch, P. Bucher, PROSITE: a documented database using patterns and profiles as motif descriptors, *Brief Bioinform.* 3 (2002) 265–274.
- [16] B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, *J. Mol. Biol.* 55 (1971) 379–400.

- [17] A. Nicholls, K.A. Sharp, B. Honig, Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins* 11 (1991) 281–296.
- [18] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol.* 157 (1982) 105–132.
- [19] J.L. Guy, R.M. Jackson, K.R. Acharya, E.D. Sturrock, N.M. Hooper, A.J. Turner, Angiotensin-converting enzyme-2 (ACE2): comparative modeling of the active site, specificity requirements, and chloride dependence, *Biochemistry* 42 (2003) 13185–13192.
- [20] O. Spiga, A. Bernini, A. Ciutti, S. Chiellini, N. Menciassi, F. Finetti, V. Causarone, F. Anselmi, F. Prischi, N. Niccolai, Molecular modelling of S1 and S2 subunits of SARS coronavirus spike glycoprotein, *Biochem. Biophys. Res. Commun.* 310 (2003) 78–83.